

Claims

- [c1] 1. A method for identifying documents most relevant to a query from a collection of documents that is organized based on a set of indices, said method comprising the steps of:
- a) determining a query class for the query, said query class associated with a routing function and a ranking function, said routing function capable of determining subsets of the collection that most likely include the most relevant documents, and said ranking function capable of sorting the documents in terms of relevancy;
 - b) identifying a set of indices most relevant to said query;
 - c) identifying a set of documents related to said query based on said determined indices, said identification performed via passing said ranking function associated with said determined query class along with said query to each search engine that manages a determined index from a collection of relevant indices;
 - d) collecting results ranked based upon said ranking function and merging and sorting said collected results by relevancy; and
 - e) returning a subset of the highest ranked documents

as the documents most relevant to the query.

[c2] 2. The method as per claim 1, wherein said step for determining a query class further comprises the following steps:

a) analyzing user profile data, user search context and history data, log file data, and index statistics, or other query related external data; and

b) utilizing said analyzed data in determining a query class for said search query.

[c3] 3. The method as per claim 1, wherein said step for identifying a set of indices further comprises the step of using routing information obtained from applying said routing function associated with said query class to determine said set of indices.

[c4] 4. The method as per claim 1, wherein said step of returning a subset of the highest ranked documents further comprises the following steps:

a) assigning each search result item a relevancy score; and

b) returning a predetermined subset of results from said search results.

[c5] 5. The method as per claim 4, wherein said method additionally comprises the step of sorting search results by

said relevancy score in decreasing order prior to returning said predetermined subset of results.

[c6] 6. A method as per claim 1, wherein said method is implemented across networks.

[c7] 7. A method as per claim 6, wherein said across networks element comprises any of, or a combination of, the following: wide area network (WAN), local area network (LAN), cellular, wireless, or the Internet.

[c8] 8. An article of manufacture comprising a computer user medium having computer readable code embodied therein which identifies documents most relevant to a query from a collection of documents that is organized based on a set of indices, said medium comprising:

- a) computer readable program code determining a query class for the query, said query class associated with a routing function and a ranking function, said routing function capable of determining subsets of the collection that most likely include the most relevant documents, and said ranking function capable of sorting the documents in terms of relevancy;
- b) computer readable program code determining indices most relevant to said query;
- c) computer readable program code identifying a set of documents related to said query based on said deter-

mined indices, said identification performed via passing said ranking function associated with said determined query class along with said query to each search engine that manages a determined index from a collection of relevant indices;

d) computer readable program code collecting results ranked based upon said ranking function and merging and sorting said collected results by relevancy; and

e) computer readable program code returning a subset of the highest ranked documents as the documents most relevant to the query.

[c9] 9. An article of manufacture as per claim 8, wherein said computer readable program code determining a query class further comprises:

a) computer readable program code analyzing user profile data, user search context and history data, log file data, and index statistics, or other query related external data; and

b) computer readable program code utilizing said analyzed data in determining a query class for said search query.

[c10] 10. An article of manufacture as per claim 8, wherein said computer readable program code identifying a set of indices further comprises computer readable program code using routing information obtained from applying

said routing function associated with said query class to determine said set of indices.

[c11] 11. An article of manufacture as per claim 8, wherein said computer readable program code returning a subset of the highest ranked documents further comprises:

- a) computer readable program code assigning each search result item a normalized score;
- b) computer readable program code sorting search results by score in decreasing order of said scores; and
- d) computer readable program code returning a predetermined subset of results from said sorted list of search results.

[c12] 12. A method for retrieving information comprising the steps of:

- a) receiving a query;
- b) parsing said query and generating a set of query terms;
- c) identifying statistical information regarding each of said query terms and different permutations of query terms;
- d) identifying lexical affinities associated with said permutations of query terms;
- e) classifying said query into a query category based upon results of steps c and d;
- f) identifying a set of ranking parameters associated with

said query category;
g) identifying routing information associated with said query category;
h) issuing a query to a search engine by applying said identified ranking parameters and said identified routing information; and
i) receiving and rendering search results from said search engine.

[c13] 13. A method as per claim 12, wherein said step of identifying statistical information additionally comprises the step of analyzing log data.

[c14] 14. A method as per claim 12, wherein said step of identifying statistical information additionally comprises the step of analyzing user feedback.

[c15] 15. A method as per claim 12, wherein said method is implemented across networks.

[c16] 16. A method as per claim 15, wherein said across networks element comprises any of, or a combination of, the following: wide area network (WAN), local area network (LAN), cellular, wireless, or the Internet.

[c17] 17. An article of manufacture comprising a computer user medium having computer readable code embodied therein for retrieving information comprising the steps

of:

- a) computer readable program code receiving a query;
- b) computer readable program code parsing said query and generating a set of query terms;
- c) computer readable program code identifying statistical information regarding each of said query terms and different permutations of query terms;
- d) computer readable program code identifying lexical affinities associated with said permutations of query terms;
- e) computer readable program code classifying said query into a query category based upon results of steps c and d;
- f) computer readable program code identifying a set of ranking parameters associated with said query category;
- g) computer readable program code identifying routing information associated with said query category;
- h) computer readable program code issuing a query to a search engine by applying said identified ranking parameters and said identified routing information; and
- i) computer readable program code receiving and rendering search results from said search engine.